А.І. Багдзевіч, А.Ю. Станкевіч (Гродно, ГрГУ)

ПРАЕКТАВАННЕ ЛІНГВІСТЫЧНАЙ БАЗЫ ДАДЗЕНЫХ «МАРФАНАЛАГІЧНАЕ ВЫВУЧЭННЕ ЎКРАІНСКАЙ МОВЫ»

Для збірання матэрыялу па вывучэнні марфаналагічных асаблівасцей украінскай мовы для далейшага супастаўляльнага вывучэння марфаналогіі ўсходнеславянскіх моў намі была распрацавана лінгвістычная база дадзеных (далей — ЛБД) «Марфаналагічнае вывучэнне ўкраінскай мовы».

Адэкватнае апісанне прадметнай вобласці ажыццяўляецца праз выяўленне найбольш рэлевантных для вывучэння марфаналагічных асаблівасцей мовы звестак і патрабуе папярэдняга вывучэння адметнасцей марфалогіі ўкраінскай мовы (пералік тыпаў скланення з вызначэннем крытэрыяў аднясення назоўнікаў да кожнага з іх, вызначэнне асноўных тыпаў асновы і інш.).

Знаёмства з граматычнай інфармацыяй [1] дазволіла нам пабудаваць наступную інфалагічную мадэль ЛБД:

На этапе праектавання ЛБД намі была складзена агульнае апісанне яе асноўных палёў.

№	Назва поля	Фармат поля	Асаблівасці запаўнення
1	Лексема	Тэкставы	Капіраванне / ўстаўка з онлайн-слоўніка
2	Аснова	Тэкставы	Капіраванне / ўстаўка з онлайн-слоўніка, рэдагаванне запісу шляхам сцірання флексій, дадавання патрэбных літар
3	Адварот	Тэкставы	Аўтазапаўненне шляхам выканання спецыяльнага запыту на дадаванне са зваротам да тэкставай функцыі StrReverse()
4	Род	Тэкставы	Выбар з падстановачнага спісу
5	Тып асновы	Тэкставы	Выбар з падстановачнага спісу
6	Канчатак	Тэкставы	Запаўненне ўручную шляхам уводу адпаведнай літары / літар у поле з маскай «-»<СССССС
7	Тып скланення	Лічбавы	Выбар з падстановачнага спісу
8	Акцэнтны тып	Тэкставы	Выбар з падстановачнага спісу
9	Парадыгма	Поле MEMO	Капіраванне / ўстаўка з онлайн-слоўніка

Далей быў здзейснены пераход да СКБД Microsoft Access 2007.

У якасці крыніцы матэрыялу быў абраны онлайн-слоўнік украінскай мовы з «Украінскага лінгвістычнага парталу» [2], які мае 256 618 уваходжанняў. Матэрыял для запаўнення ЛБД адбіраўся метадам суцэльнай выбаркі субстантыўных лексем з дадзенага слоўніка.

Дадзеныя, што ўносіліся ў ЛБД, уключаюць у сябе: а) моўны матэрыял са слоўніка (лексема, парадыгма); б) лінгвістычны матэрыял са слоўніка (звесткі аб родзе лексемы); в) лінгвістычны матэрыял, які ўносіўся самастойна (інфармацыя аб тыпе скланення, тыпе асновы, акцэнтным тыпе); г) моўны матэрыял, які ўносіўся самастойна (аснова); д) матэрыял тэхнічнага характару, які генерыраваўся аўтаматычна (адва-

рот). Для вывучэння марфаналогіі субстантыўнай асновы надзвычай важна групаваць дадзеныя ў алфавітным парадку па зыходзе асновы. Менавіта з гэтай мэтай у структуру ЛБД было ўключана поле [Адварот], якое запаўняецца праз запыт на абнаўленне па наступнай інструкцыі SQL: UPDATE [Украінская мова] SET [Украінская мова]. Адварот = StrReverse([Аснова]);

Лексема	**Аснова**	**Адварот**
аба́зія	аба́зій	йіз́аба
аба́к	аба́к	ќаба
абази́нка	абази́нк	кни́заба
абази́нець	абази́нец'	'цени́заба

Выкананне запыту «Адварот» адбываецца пасля ўвядзення пэўнай колькасці дадзеных у ЛБД пры ўзнікненні неабходнасці сартавання матэрыялу.

Напаўненне ЛБД — працэс працаёмкі і часаёмкі. Але для зручнасці, павышэння эфектыўнасці і нагляднасці, памяншэння часа- і працаёмкасці дадзенага працэсу намі былі распрацаваны спецыяльныя формы і алгарытм. Форма «Для Index'у» была сканструявана такім чынам, каб дазволіць уносіць у ЛБД тую частку звестак, якая бярэцца непасрэдна са слоўніка або фармуецца на падставе дадзеных слоўніка. Так, палі [Канчатак], [Тып асновы], [Адварот] у дадзеную форму не ўключаліся, паколькі іх запаўненне больш эфектыўнае ў рэжыме табліцы або ўвогуле ажыццяўляецца аўтаматычна.

Запаўненне палёў адбывалася ў такім парадку: [Лексема] — [Аснова] — [Тып скланення] — [Парадыгма]. Пасля на падставе дадзеных з апошняга поля запаўняліся таксама звесткі аб акцэнтным тыпе. Поле [Род] запаўнялася толькі ў тым выпадку, калі было бачна, што самастойнае вызначэнне роду без дапамогі слоўніка выкліча цяжкасці, у астатніх выпадках гэтае поле на этапе запаўнення формы заставалася пустым, бо яго больш хутка і эфектыўна можна запаўняць у рэжыме табліцы.

Алгарытм напаўнення ЛБД выглядаў наступным чынам:

1. Запаўненне формы «Для Index'у».
   a. Унясенне лексемы.
   b. Унясенне асновы (шляхам ручной карэкцыі).
   c. Выстаўленне роду ў полі [Род] пры наяўнасці сумненняў у магчымасці самастойнага вызначэння (у адваротным выпадку гэтая аперацыя пераносіцца на этап 3).
   d. Выстаўленне тыпу скланення.
   e. Унясенне парадыгмы.
   f. Выстаўленне акцэнтнага тыпу.
2. Аўтаматычнае запаўненне палёў і генерацыя табліц на падставе наяўных дадзеных.
   a. Выкананне запыту «Адварот».
   b. Генерацыя табліцы «Парадыгма».
3. Запаўненне ЛБД у таблічным рэжыме.

a. Выстаўленне роду (калі не запоўнены ў рэжыме формы, гл. п. 1 с.). Адбываецца паэтапна: м.р., ж.р., н.р. Пасля запаўнення поля для лексем аднаго роду з дапамогай фільтра хаваюцца запоўненыя радкі і ўносяцца дадзеныя для лексем наступнага роду.

b. Выстаўленне канчатка. Ажыццяўляецца для зручнасці тым самым спосабам, што і п. 2 а.

c. Выстаўленне тыпу асновы. Ажыццяўляецца шляхам хавання радкоў з лексемамі 3, 4 тыпаў скланення, а таксама нескланяльных і іншых тыпаў (напрыклад, ад’ектыўнага), з папярэднім алфавітным сартаваннем радкоў паводле поля [Адварот], што значна паскарае працэс вызначэння тыпу асновы.

4. Запаўненне формы «Марфаналагічныя з’явы».

Каментарый да п. 2b. Генерацыя табліцы «Парадыгма» з палямі [Код], [Парадыгма] і палямі для кожнага з 7 склонаў выконваецца са зваротам да дадзеных табліцы «Украінская мова» праз серыю інструкцый SQL.

Разбіўка поля [Парадыгма] на палі, якія змяшчаюць значэнні кожнага склону, ажыццяўляецца праз зварот да тэкставых функцый Instr(), Len(), Mid() і праводзіцца па пазнаках імёнаў склонаў.

Пазначым праз L і R імёны першага (левага) і другога (правага) кампанентаў пары склонаў з такога мноства пар: {(«називний», «родовий»), («родовий», «давальний»), («давальний», «знахідний»), («знахідний», «орудний»), («орудний», «місцевий»), («місцевий», «кличний»)}. У агульным выпадку частка інструкцыі SQL, якая адказвае за запаўненне палёў са значэннем склону L такая: Mid([Парадыгма], InStr([Парадыгма],L)+Len(L), InStr([Парадыгма],R)-InStr([Парадыгма],L)-Len(L)).

Тут тэкставая функцыя Mid() прымае 3 аргументы: 1) імя радка: [Парадыгма]; 2) пазіцыя сімвала радка, з якога пачынаецца вяртаемы падрадок: InStr([Парадыгма],L)+Len(L); 3) даўжыня вяртаемага падрадка: InStr([Парадыгма],R)-InStr([Парадыгма],L)-Len(L).

Інструкцыя выконваецца толькі пры ўмове ўваходжання імёнаў склонаў L і R ў полі [Парадыгма].

У прыватным выпадку, калі імя склону функцыянуе як мяжа па-за парай L–R (імя «кличний» як мяжа для склону «кличний»; імя «називний» як мяжа для склону «називний» у няпоўных парадыгмах з 1 кампаненту), частка інструкцыі SQL, якая адказвае за запаўненне палёў для гэтага склону, такая: Mid([Парадыгма], InStr([Парадыгма],X)+Len(X)).

Тут для склона X тэкставая функцыя Mid() прымае 2 аргументы, у выніку вяртаюцца ўсе знакі ад пазіцыі, зададзенай другім аргументам, да канца радка.

Заўважым, што марфалагічныя атрыбуты слова, якія вызначаны распрацаванай інфалагічнай мадэллю, дазволяць аўтаматызавана пазначаць на радках ускосных склонавых форм уваходы чаргаванняў, усячэнняў і іншых марфаналагічных з'яў. Распрацоўка модуля такой разметкі плануецца ў перспектыве.

Каментарый да п. 4. Форма «Марфаналагічныя з’явы» была створана спецыяльна для вызначэння наяўнасці марфаналагічных з’яў у зыходзе субстантыўнай асновы. Дадзеная форма ўключае ў сябе толькі 4 палі, 3 з якіх ужо запоўненыя на папярэдніх этапах ([Лексема], [Аснова], [Парадыгма]), а запаўняецца толькі адно поле [Наяўнасць марфаналагічных з’яў]. Форма пабудавана такім чынам, каб запісы лексемы, асновы і парадыгмы былі добра візуалізаваныя, што дазваляе хуткі прагляд парадыгмы і ўнясенне ў адзінае поле, якое запаўняецца, адпаведных дадзеных. Каб існавала магчымасць выяўлення новых і нетыповых чаргаванняў і іншых марфаналагічных з’яў, дадзенае поле было зроблена тэкставым і без магчымасцей падстаноўкі.

З мэтаю вывучэння заканамернасцей марфаналагічнага ўзроўню ў ЛБД была створана папярэдняя сістэма запытаў на выбарку. Найперш цікавасць уяўляюць запыты на выбарку паводле фінала субстантыўнай асновы, якія ў той жа час адлюстроўваюць разнастайную інфармацыю аб выбраных субстантывах: род, тып скланення, тып асновы, наяўнасць чаргаванняў. Так, выбарка назоўнікаў з асновамі на [л’] і [н’] дала 119 уваходжанняў, з якіх 3 адносяцца да першага тыпу скланення, 106 — да другога, 6 — да трэцяга, 1 — да чацвёртага, 1 — нескланяльны, 1 — субстантыў ад’ектыўнага скланення. Пакажам фрагмент выніковай табліцы паводле гэтага запыту.

Запыт «Канец асновы»					
Аснова	**Тып скланення**	**Род**	**Тып асновы**	**Наяўнасць марфаналагічных з'яў**	**Адварот**
А́бел'	2	мужч.	мяккая	м/цв	'леб́А
авансува́нн'	2	ніякі	мяккая	н	'нн́авуснава
автомагістра́л'	3	жаночы	-	м/цв	'л́артсігамотва
автомайсте́рн'	1	жаночы	мяккая	м/цв	'нр́етсйамотва
автоматизува́нн'	2	ніякі	мяккая	н	'нн́авузитамотва

Запыт «Парадыгма» дазваляе вывесці ў асобную табліцу парадыгмы абранай групы / груп субстантываў. Створаны таксама запыт «Асновы змешанага тыпу 1 скланення з чаргаваннямі» (вынік — усяго 4 выпадкі на 2 200 уваходжанняў).

Як бачым, спраектаваная на аснове распрацаванай інфалагічнай мадэлі лінгвістычная база дадзеных дазваляе групаваць матэрыял, рабіць выбаркі, супастаўляць дадзеныя з улікам найбольш рэлевантных у адносінах да вывучэння марфаналагічнага падузроўню мовы аспекта.

ЛІТАРАТУРА

1. Горпинич, В.О. Сучасна українська літературна мова. Морфеміка. Словотвір. Морфонологія: навч. посіб. — Київ, 1999.

2. Словники України on-line / Український лінвгістичний портал [Электронный ресурс]. — Режим доступа: http://lcorp.ulif.org.ua/dictua.

3. Рудикова, Л.В. Проектирование баз данных. — Минск, 2009.